NASA-RP-1048 19790025687

NASA Reference Publication 1048

Procedure for Extraction of Disparate Data From Maps Into Computerized Data Bases

Bobby G. Junkin

OCTOBER 1979

LIBRARY COPY

OCT 30 1979

LANGLEY RESEARCH CENTER
LIBRARY, NASA
HAMPTON, VIRGINIA

NASA

NASA Reference Publication 1048

Procedure for Extraction of Disparate Data From Maps Into Computerized Data Bases

Bobby G. Junkin
Earth Resources Laboratory
NSTL Station, Mississippi

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1979

Contents

Table

Figures

Procedure for Extraction of Disparate Data From Maps Into Computerized Data Bases

SUMMARY

This report presents a procedure for the extraction of disparate sources of data from geographic maps and the subsequent conversion of these data into a format that is suitable for processing on a computer-oriented information system.

Also presented are several graphic digitizing considerations that are important from the standpoint of computer processing, map sources and characteristics, digitizing equipment, and data structuring methods. These considerations, as they relate to the NASA Earth Resources Laboratory's Digitizer System, are intended to serve as guides to those organizations contemplating the conversion of map-based data using any comparable map digitizing system. Current operating procedures for the NASA Earth Resources Laboratory's Digitizer System are presented in a simplified and logical manner for easy use by organizations with digitizing requirements.

INTRODUCTION

Organizations that are considering the use of map-based data to derive information for management decisions are eventually faced with the problem of converting disparate sources of data from a map format to a form that is suitable for processing on a computer-oriented information system. In converting data from map sources—frequently referred to as "digitizing"—there are several considerations that are important from the standpoint of computer processing, map sources and characteristics, digitizing equipment, and data structuring methods. This report covers these considerations as they relate to the NASA Earth Resources Laboratory's (ERL) Digitizer System. It is intended to serve as a guide to organizations contemplating conversion of map-based data using any comparable map digitizing system.

This report also presents the documentation of operating procedures for the NASA ERL Digitizer System. Because this system involves the conversion of geographical map data into digital format for further computer processing, it is appropriate to consider where this system falls within the spectrum of computer-oriented information systems. In addition, it is necessary to consider what digitizing consists of, the role of maps, and one concept for a computer-oriented information system. Several of these concepts, because of their tutorial nature, are relegated to the appendixes. (See appendix A entitled "Geographic Map Sources, Characteristics, and Classifications" and appendix B entitled "Graphic Digitizing.") It is assumed that the user of this report has a basic background in computers and, in addition, basic knowledge relating to graphic map digitizing.

In compliance with the NASA's publication policy, the original units of measure have been converted to the equivalent value in the Systeme International d'Unites (SI). As an aid to the reader, the SI units are written first and the original units are written parenthetically thereafter.

COMPUTER-ORIENTED INFORMATION SYSTEMS CONCEPT

There is a recognized need on the part of management within the federal, state, and regional sectors of the government—as well as in the private sector—for current information relevant to land management decisions and for evaluating the alternatives in land use and the subsequent effects of resource development on the environment. This information is acquired by compiling detailed data bases from disparate, geographically oriented base maps or from remote sensor data obtained from aircraft and satellites.[1] These data bases contain information on land use, elevation, slope, soil series, rainfall, population density, etc. The capability to manipulate, store, analyze, display, and disseminate the large volumes of data in these data bases is provided with a computer-oriented information system. A block diagram for one concept of such a system is shown in figure 1.

The graphic map data are digitized and input to the computer by either a magnetic tape or an online process. Tabular data are input by such devices as graphic terminals, card readers, or paper tape punches. Multispectral scanner imagery data are input via computer-compatible tapes, with user inputs being entered via an interactive image processing system. Subsequent processing within the mainframe can vary from simple procedures (e.g., input data display for screening) to very sophisticated analysis procedures (e.g., supervised or unsupervised spectral-pattern-recognition classification techniques or processing through application algorithms). The processed data that are output are presented on a line printer, a CRT display, a card punch, an electrostatic printer-plotter, or a color strip film recorder.

A system closely resembling this concept has evolved at the ERL through the research efforts documented in references 1 and 2. Through these efforts, the ERL has developed a computerized land resources information system capability to automatically generate color-coded land use maps of very large areas through the use of multispectral scanner data, sensed remotely from either spacecraft or aircraft platforms. Basically, this capability provides potential users of multispectral scanner data with a procedure, and the associated hardware and software, for producing land use maps by automated computerized systems. The steps in the ERL procedure for generation of land use maps from multispectral scanner data involve (1) task planning, (2) data acquisition, and (3) data processing. Task planning includes definition of the potential application; development of survey requirements; selection of platform, sensors, and data bands; and requesting or acquiring data. Data acquisition is accomplished through the use of remote sensing from space or aircraft platforms. The procedures for data processing include, but are not limited to, the following: (1) preprocess data, (2) screen and evaluate data, (3) prepare display tapes, (4) select and extract training samples, (5) analyze training samples statistically, (6) determine best bands for pattern recognition classification, (7) generate classification tables, (8) extract best bands from total set, (9) classify unknown data sets, (10) color-code, scale, and rectify land use map, (11) record land use map on color film, (12) prepare scorecard, and (13) prepare final product. Examples of the output products are standard statistical variables; acreage compilation by class of material; and a scaled, color-coded land use map as continuous strips of film or color paper. The ERL data analysis system (DAS), for processing remote-sensor data, has the capability to accept multispectral remote-sensor data in computer-compatible nine-track or analog magnetic tape format or from photographic film (either black and white or color transparencies).

Examples of several practical, modular systems, with emphasis on low cost, are given in reference 2. These systems, an example of which is shown in figure 2, consist of an image display system, a graphic digitizer, a small digital computer, and an output recording device. All hardware components used in these low-cost data processing systems are off-the-shelf. The software consists of a Landsat multispectral scanner data reformatting program, a series of supervised and unsupervised spectral-pattern-recognition programs, a program to reference the image data to a map base, a data storage and retrieval program, and various applications programs. These software programs are available through COSMIC, NASA's software dissemination facility. (They can be obtained by contacting the University of Georgia, Athens, Georgia.)

NASA ERL DIGITIZER SYSTEM

Discussion

The NASA ERL Digitizer System is an integral part of the computer-oriented information system described in the previous section. The digitizer has been added to the systems specifically detailed in references 1 and 2. As such, this general purpose electronic digitizer system has the capability to con-

[1]Some map data, e.g., National Cartographic Information Center (NCIC) digital terrain tapes, are available in a digitized form (see appendix A).

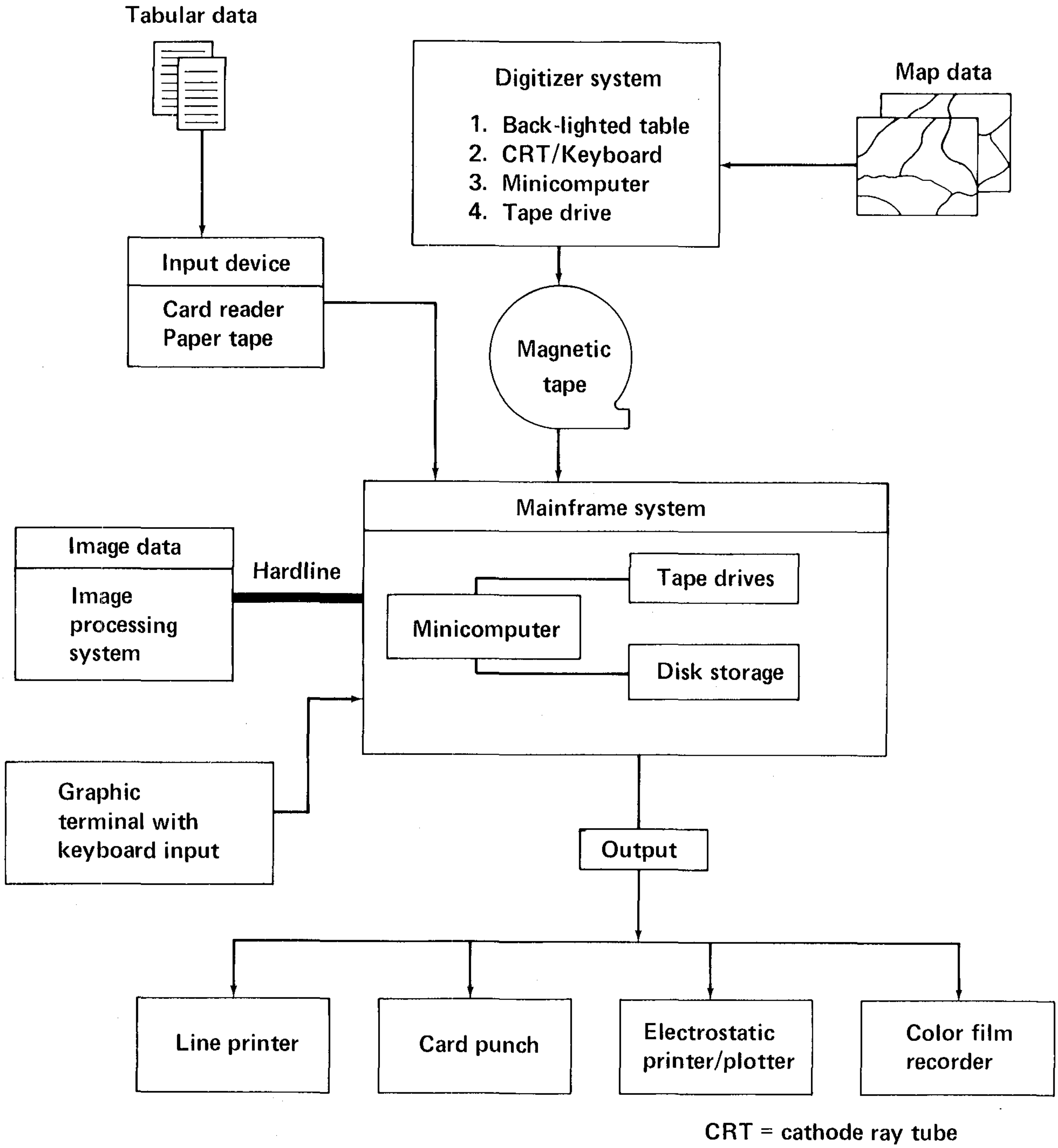
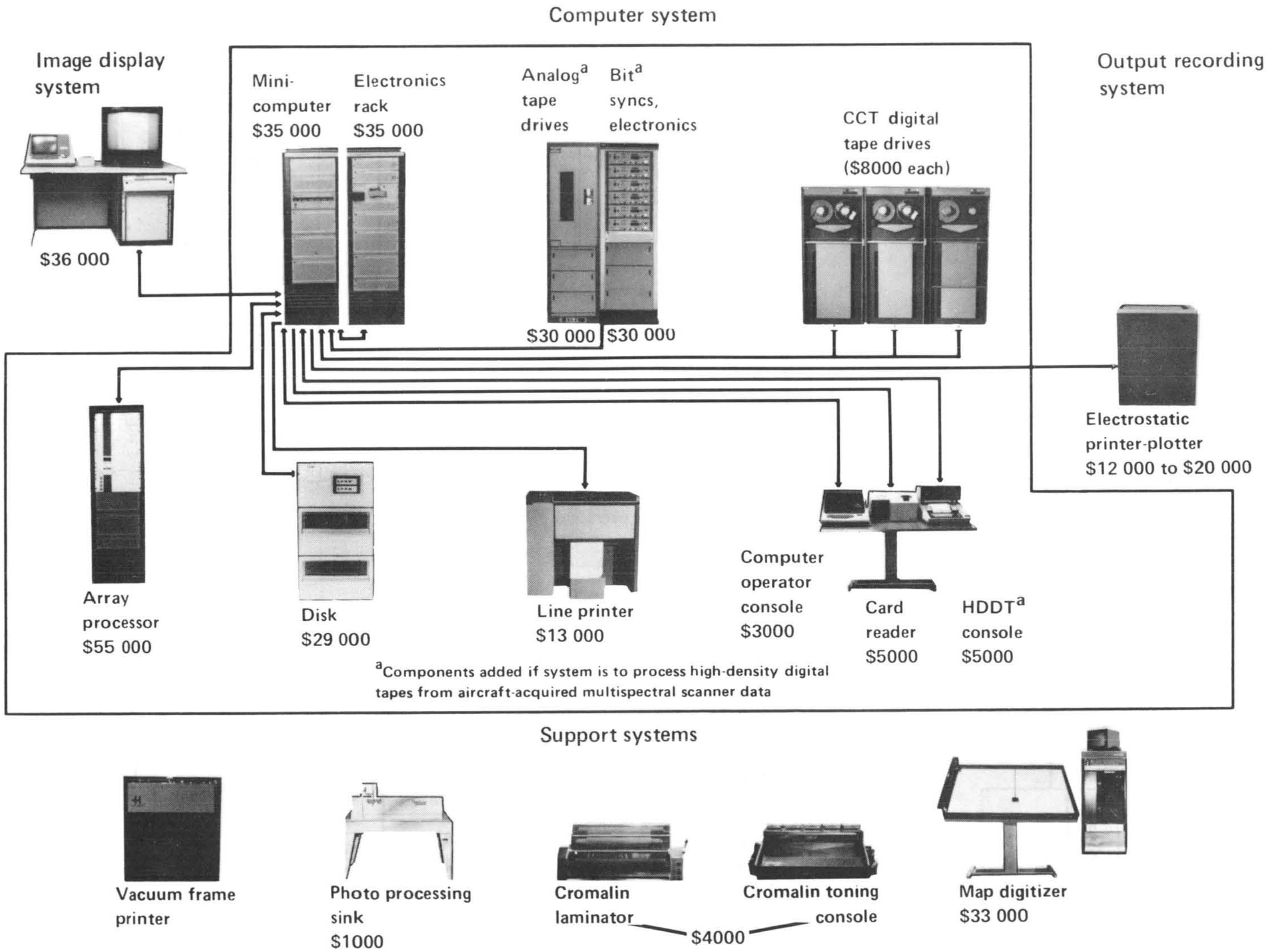


FIGURE 1.—General concept of computer-oriented information system.

vert large volumes of data contained on maps, charts, and film into a digital format for further processing on a mainframe computer system such as that described in references 1 and 2. It is semi-automated in the sense that an operator is required to perform certain manual operations when the digitizer is being operated in either the point or line mode of data capture.



NOTES: CCT = computer-compatible tape
HDDT = high-density digital tape

FIGURE 2.—Example of low-cost data analysis system.

System Components

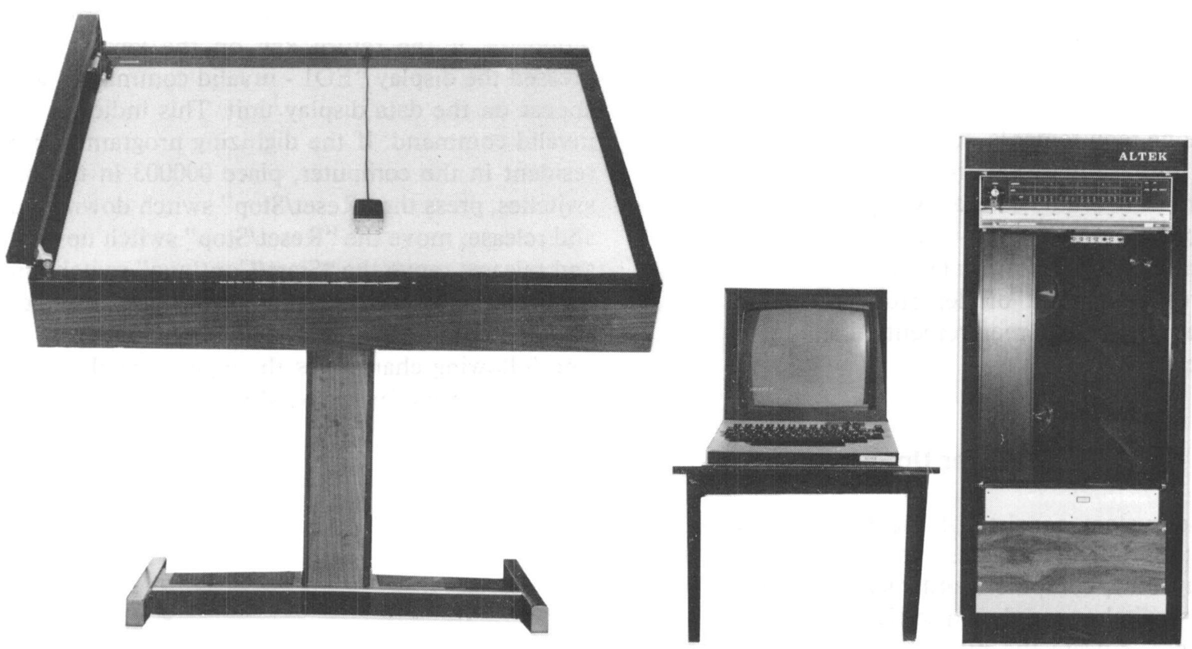
Components of the system are shown in figure 3. These components were manufactured by various companies and assembled by the Altek Corporation. Several other companies provide similar systems at competitive costs.

The computer processor is a Data General Corporation NOVA 1210 series. It currently contains a core memory of 8000 words (16-bit words). The nine-track, 800 BPI density magnetic tape unit is the output storage medium for the digitized data. Manual data entry is accomplished via the keyboard. The capability to tag data, enter data qualifiers, etc., is provided through the use of the numbers 0 to 9 and 18 alphanumeric characters. Data input via the keyboard or actual digitized data from the source can be displayed on the CRT. The maximum area for digitizing on the back-lighted table is approximately 91 by 106 centimeters (36 by 42 inches). Table lighting is primarily for use in reading information from transparencies, and the intensity of the light can be controlled by the operator. There are also tilt and height adjustments for the table. Data recording is accomplished by using a free moving, handheld device called a cursor. This cursor has a viewing window 5 centimeters (2 inches) in diameter with crosshairs for reading precise points or tracing lines. Pushbuttons are provided on the cursor for data recording.

OPERATING PROCEDURES FOR THE NASA ERL DIGITIZER SYSTEM

Discussion

The NASA ERL Digitizer System, which is modular in design, consists of components that are manufactured by various companies. Each component has technical manuals that contain specifications and basic operating procedures. A single manual entitled "Part I DATALAB 'IMP' Interactive Mapping Program" was written by Altek programers and describes all functions present in the system. This manual, however, does not provide detailed step-by-step procedures that would allow digitizing by a person who is not intimately familiar with the system. The procedures herein organize



Back-lighted table CRT NOVA 1210 computer and magnetic tape unit

FIGURE 3.—Components of the NASA ERL digitizer system.

the functions, commands, and steps in a simplified, logical manner for use by an individual with a requirement to digitize data.

The system has some features that each user should understand. Basically, the system senses a very fine grid of wires in the tabletop via the cursor, converts these impulses into digital form, and stores the digital data on a magnetic tape. The grid is sensed in a true X-Y matrix. The smallest distance between discrete data points measurable by the system is 0.0025 centimeter (0.001 inch). By applying orientation and scale factors to this basic measurement, almost any kind of data can be digitized, stored, and processed in real numbers rather than language or numbers unique to the digitizer. A variety of formats is available for storing data on magnetic tape. Also, there are three variable data strings that include a free text area available for use during the data collection operation.

The system has two specific modes of operation—point mode and line mode. In the line mode, data points are taken each time the cursor is moved a specified distance by holding any of the cursor buttons down. In the point mode, data are taken each time a cursor button is pressed. As implied, the line mode is intended to be used when collecting linear data, and the point mode is intended for taking individual points. Singular points may also be taken while in the line mode.

These procedures, which are applicable to most digitizing requirements, set up guidelines for taking data from source material with a true X-Y grid system: X coordinates in a six-digit integer number, Y coordinates in a seven-digit integer number, and alphanumeric labeling data. By using these procedures as a point of departure, procedures for unique digitizing requirements can readily be written.

Power Up

The system consists of the following components: back-lighted table, cursor, keyboard, data display unit, computer, and magnetic tape drive. Power switches are shown in figures 4 to 6.

Step 1. Locate the main power switch for the back-lighted table (fig. 4) and move upward to the "on" position.

Step 2. Locate the tabletop switch (fig. 4) and rotate clockwise to the "on" position. The lamp intensity can be controlled by rotating this switch.

Step 3. Locate the power key for the computer (fig. 5) and rotate the key clockwise to the word "lock." When lights are observed in the address and data areas, rotate the key counter-clockwise to the word "on."

Step 4. Locate the power switch for the tape drive (fig. 5) and press. The green light to the left of this switch and the word "power" at the bottom of the unit will illuminate.

Step 5. Set duplex transmission switch on the back of the data display unit to "full" position.

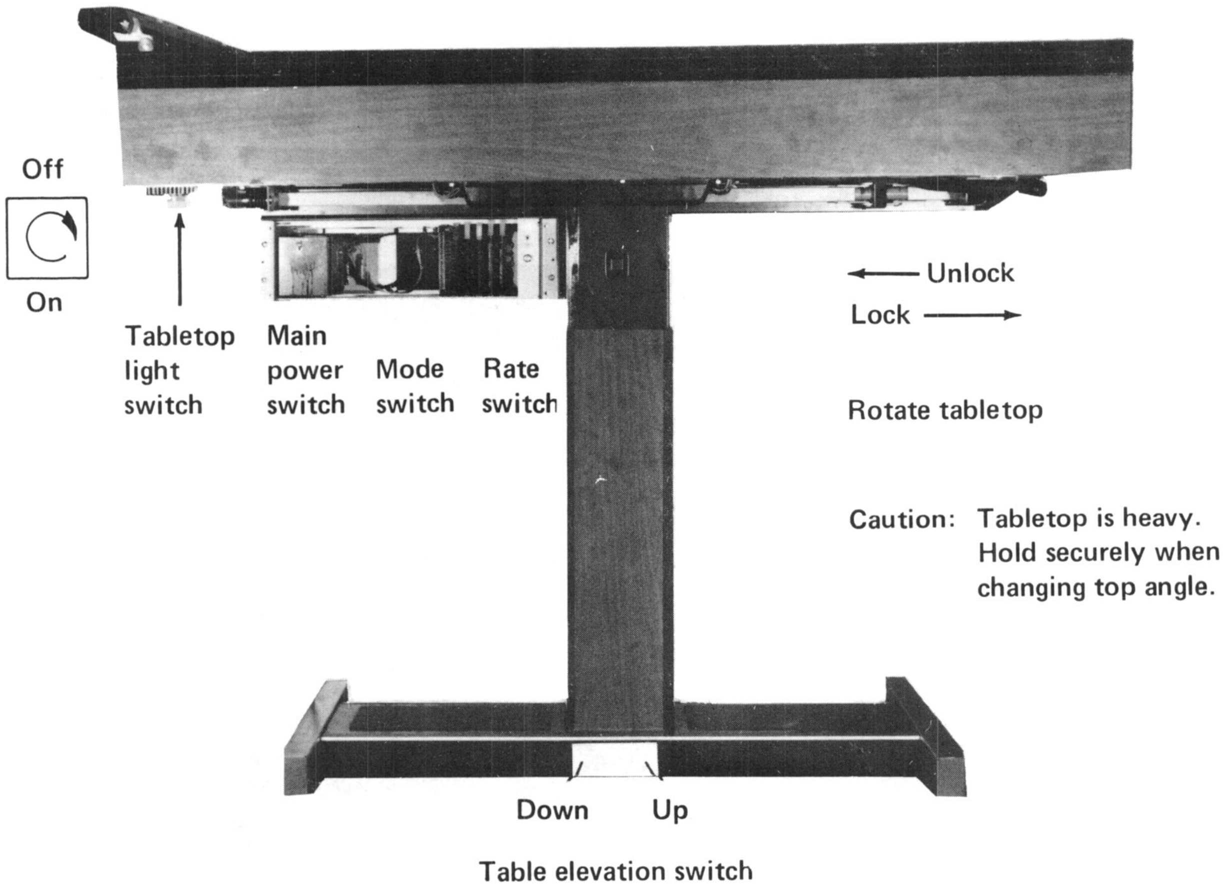
Step 6. Locate the power switch on the back of the data display unit (fig. 6) and press the "on" position.

Program Loading

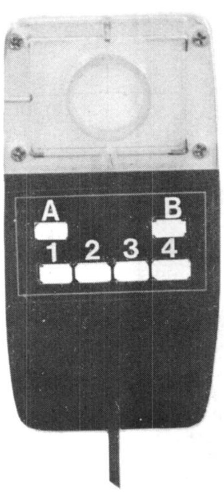
The digitizing program will normally be resident in the computer even if the power has been off. An asterisk (*) will appear on the data display unit in addition to an X and Y coordinate if the program is resident. When the cursor is moved, the X and Y coordinates change if the program is resident. Additionally, if the return key on the keyboard is pressed the display "EO1 - invalid command" will appear on the data display unit. This indicates an invalid command. If the digitizing program is not resident in the computer, place 000003 in the bit switches, press the "Reset/Stop" switch downward and release; move the "Reset/Stop" switch upward and release; move the "Start/Continue" switch upward and release. The setting of the octal number 000003 in the bit switches is illustrated in figure 7. The following chart gives the equivalent three-bit binary numbers for the octal numbers 0 to 7.

Binary	*Octal*
000	0
001	1
010	2
011	3
100	4
101	5
110	6
111	7

Back-lighted table



Cursor



Keyboard

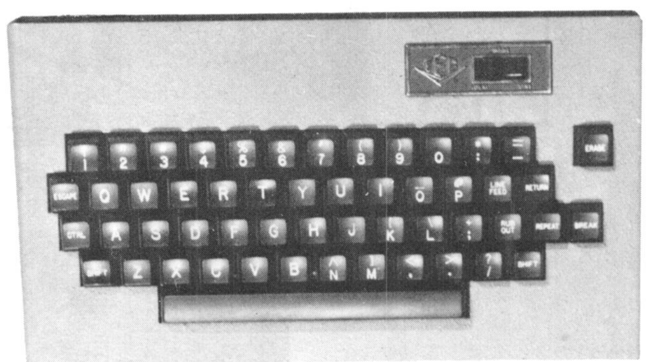


FIGURE 4.—Back-lighted table, keyboard, and cursor.

If this procedure fails, go to the following program-loading schedule.

Step 1. Mount the program tape without a write ring on the lower tape drive spool, threading the tape as shown on the tape drive.

Step 2. Press the "load" button.

Step 3. After the tape has stopped moving, press the "on line" button—the green light opposite this button will illuminate.

Step 4. Follow steps 1 to 24 in table I.

Step 5. At this point, the program tape is properly mounted, the bootstrap program is loaded into

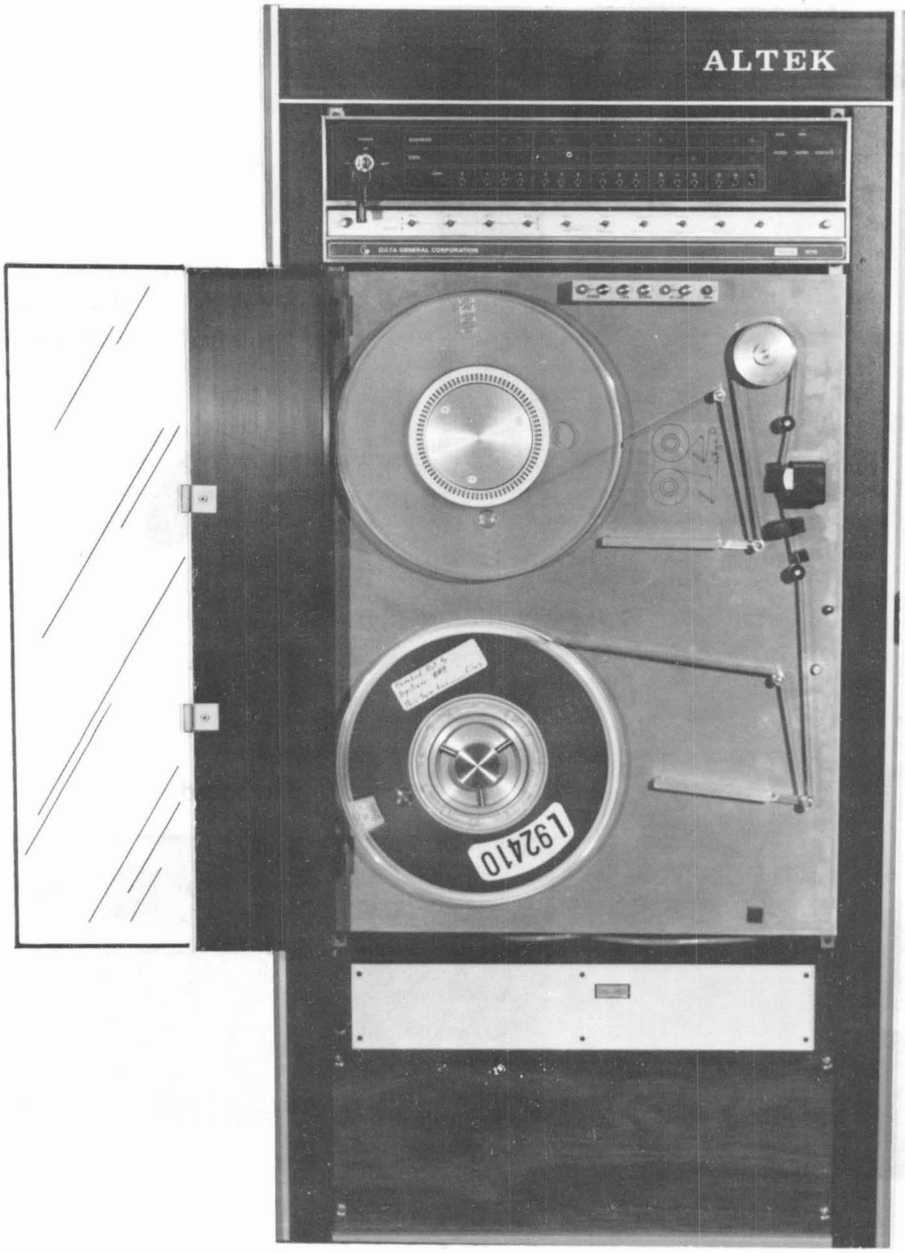
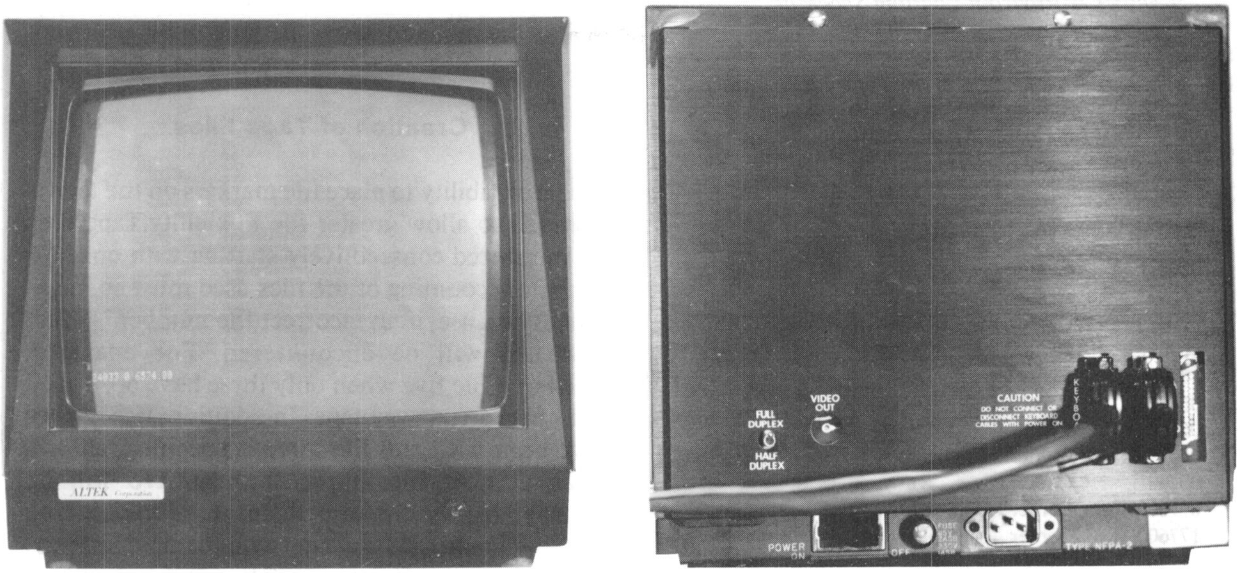


FIGURE 5.—NOVA 1210 computer and magnetic tape unit.



Front　　　　Back

FIGURE 6.—Front and back views of data display unit.

the computer, and the digitizing program can be loaded from the tape into the computer. Press the "Reset/Stop" switch downward and release to stop the computer.

Step 6. Move the "Reset/Stop" switch upward and release to set the computer at the proper memory location.

Step 7. Move the "Start/Continue" switch upward and release to start the computer operating again. The program is loaded from tape, and an asterisk will appear on the data display unit.

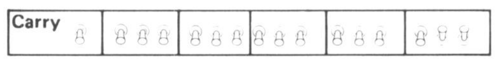
Step 8. Set 000003 in the bit switches. This will allow for the clearance of most computer stoppages that will be encountered by executing "Stop," "Reset," and "Start" in sequence.

Step 9. Rewind the program tape by pressing the tape drive "Rewind" switch, and when the tape stops press the "Rewind" switch again.

Step 10. Remove the program tape, and store it away from heat and magnetic influence.

Tape Mounting

The processor buffer will hold only nine data points; therefore, a tape must be mounted if more than nine points are to be read. Magnetic tapes on which data will be recorded must have a "write" ring installed. When mounting the program tape, however, ensure that a "write" ring is *not installed*. This will prevent inadvertent destruction of the program tape.



Decimal	0	0	0	0	0	3
Binary	000	000	000	000	000	011

FIGURE 7.—Example of bit switch setting.

Step 1. Pull out the center knob on the supply spindle approximately 1.25 centimeters (0.5 inch) allowing easy positioning of the tape reel (fig. 5).

Step 2. Place the reel of tape to be mounted on the supply spindle, ensuring that the reel is in the full rear position. Push the center knob until a click is heard.

Step 3. Thread the tape as shown by the decal on the tape drive, ensuring that there is no slack in the

TABLE I.—Program Loading Schedule

Step	Set in bit switches	Position of Examine/ Examine Next switch	Position of Deposit/ Deposit Next switch	Data register should read (a)
1	017760	Up	Center	
2	020411	Center	Up	
3	062022	Center	Down	
4	020410	Center	Down	
5	063022	Center	Down	
6	102400	Center	Down	
7	061122	Center	Down	
8	063622	Center	Down	
9	000777	Center	Down	
10	002401	Center	Down	
11	017600	Center	Down	
12	177600	Center	Down	
13	017760	Up		020411
14	017760	Down		062022
15	017760	Down		020410
16	017760	Down		063022
17	017760	Down		102400
18	017760	Down		061122
19	017760	Down		063622
20	017760	Down		000777
21	017760	Down		002401
22	017760	Down		017600
23	017760	Down		177600
24	017760	Up		020411

[a]An incorrect value in the data register, detected during steps 13 through 24, is cause to go to step 1 and start again.

tape so that friction will cause the tape to wind onto the takeup reel.

Step 4. Press the "load" switch. The tape will wind from the supply spool to the takeup spool until the tape load point is sensed by the tape head, at which time the tape will stop.

Step 5. Press the "on line" switch. The green light immediately to the left of the "on line" switch will illuminate.

Tape Demounting

The tape must be rewound onto the supply spool to demount.

Step 1. Press the tape drive switch labeled "Rewind," and when the tape stops press the "Rewind" switch again.

Step 2. Pull out the center knob on the supply spindle approximately 1.25 centimeters (0.5 inch) to free the reel and allow the tape to be removed.

Creation of Tape Files

The capability to place file markers on the tape is provided to allow greater file flexibility.Tape files are numbered consecutively starting with one. An accurate accounting of the files used must be maintained; because, if an incorrect file number is used, problems will be encountered. For example, specifying file five when only three have been used will cause a runaway tape. In addition, if five files have been used and file three is specified, all previous data in files three through five will be destroyed. Timely and accurate bookkeeping should be adhered to during a digitizing operation. There are three situations relative to file numbers that may be encountered during a digitizing operation: (1) a new tape with no files established, (2) a used tape with known file numbers that contain good data, and (3) a used tape with good data and unknown file numbers. These situations may be handled by either of the following steps.

Step 1. Enter **MTP:1**. Press the return key, and make a note of the file number.

Step 2. Enter **MTP:** and the next higher file number. Press the return key, and make a note of the file number. Enter **TXT: find next open file** and press the return key. The system will count end-of-file markers until one lower than the number specified in this step is found and write the **TXT** statement in the first record. This is necessary because, when the **CLS** command is given in the section "Creation of Data File," step 1, the buffer is cleared, written to tape, and an end-of-file mark written where the tape head is currently positioned. If the tape head is not positioned in the first unused file, good data will be destroyed.

Mounting of Material to be Digitized

Some preparatory work is necessary before digitizing can begin. Because the system senses an X-Y grid and has the capability to convert distances to a desired scale, some known factors must be identified on the source from which data are to be recorded. Specifically, the rotation X and Y coordinate line (which should be centered in the area from which data will be taken) should be identified. In addition, the X and Y coordinate lines at the top, bottom, left, and right edges of the source material should be very accurately drawn and identified. Care should be exercised when these lines are drawn, because any error at this point is magnified throughout the data collection operation.

The grid that is sensed by the cursor is located in an area approximately 7.5 centimeters (3 inches) from the blue metal frame around the tabletop. Therefore, maps, film positives, drawings, or source material for data collection should be placed approximately 7.5 centimeters (3 inches) above the bottom metal frame and more or less centered otherwise. Tape the material to the table surface where necessary to obtain a flat working surface. It should be noted that material which has been folded, wrinkled, or otherwise mutilated will give inaccurate results when data are collected via the digitizer.

Computation of Scale Factors

The most accurate data can be collected via the digitizer by computing a scale factor in both the X and Y directions for each sheet or item of source material used. Assuming the X and Y gridlines specified in the previous section have been drawn, proceed with the following steps.

Step 1. Locate the Line/Point mode switch shown in figure 4, and move the switch upward to the point mode position.

Step 2. Enter **EVT:1** and press the return key.

Step 3. Enter **SAXS:0, 0, 1, 1** and press the return key.

Step 4. Place the cursor on the lower left X-Y grid intersection and press any cursor button. This is the point of origin.

Step 5. Place the cursor on the same X gridline at the X-Y grid intersection to the far right of the source and, without moving the cursor, press any cursor button twice.

Step 6. Find the grid intersection used in step 4, and follow the vertical gridline upward to the last X-Y grid intersection in the upper left corner. Place the cursor on this grid intersection and press any cursor button.

Step 7. Locate on the display device the X coordinate value obtained in step 5. This is the 0.0025 centimeter (0.001 inch) required to cover the number of units from the point of origin (step 4) and the grid intersection selected and identified as event 0001 in step 5.

Step 8. Divide the number of units covered from the point of origin, which is a grid intersection, to the X-Y grid intersection used when event 0001 was recorded in step 5, by the X coordinate value obtained in step 5 and shown on the display device as event 0001. The result is the X-scale factor, and it should be recorded by the operator.

Step 9. Divide the number of units covered from the point of origin to the X-Y grid intersection used when event 0002 was recorded in step 6, by the Y coordinate value obtained in step 6 and shown on the display device as event 0002. The result is the Y-scale factor, and it should be recorded by the operator.

Rotation and Initialization

In a sense, all work performed to this point has been in preparation for the interface between the source data and the digitizing system. To effect this interface, it is necessary to rotate the system so that the data points sensed by the cursor conform with the X-Y grid on the source. The following procedure accomplishes this rotation.

Step 1. Locate the X-Y coordinate intersection selected and drawn in the section "Mounting of Material to be Digitized." This point will be near the center of the area to be digitized.

Step 2. Enter **SAXS:** C coordinate value from step 1; the Y coordinate value from step 1; the X-scale value computed in the previous section, step 8; and the Y coordinate value computed in the previous section, step 9. Press the return key.

Step 3. Place the cursor on the grid intersection located in step 1, and press any cursor button.

Step 4. From the grid intersection located in step 1, follow the X gridline to the extreme right edge of the source, position the cursor on this line, and press any cursor button.

Step 5. Locate a grid intersection in each of the four corners of the source to be digitized. Place the cursor on each grid intersection selected, and check the last X-Y coordinates displayed on the screen against those under the cursor. If topographic maps are used as the source, the difference should be no greater than 0.0508 centimeter (0.02 inch) at map scale. If something other than topographic maps are used as the source, the acceptable error should be established.

Step 6. Assuming that the differences checked in step 5 are within acceptable limits, enter **EVT:1** and press the return key. If the differences are not within acceptable limits, the error probably is in scale factor computation or rotation and initialization. In this case, go back to the section on "Computation of Scale Factors" and proceed to this point.

Setting of Digitizer Output Format

The preprogramed format is 1F, 7X, 7Y, 7Z, 4E, 32C. This format is automatically set each time the program is loaded, each time a restart through the bit switches is accomplished, or when the commands **RST** and **CLR** are given. To work in actual Universal Transverse Mercator (UTM) coordinates, it is necessary to enter a format other than the preprogramed one. This is done via the **FMT:** command. Before prescribing a format, some explanation of the fields within the format is necessary. The first field, identified by the "F," consists of one space. The data loaded into this space is the cursor button used. The button and character loaded on the tape is as follows: A = blank, B = blank, 1 = 1, 2 = 2, 3 = 3 and 4 = 4. The second and third fields are the X and Y coordinate fields, respectively. These fields are identified by the "X" and "Y," and each contains 10 spaces. These spaces are occupied by a ± sign, a decimal, and eight numbers. The fourth field is a free text alphanumeric area consisting of eight spaces. This area can be changed at any time during the digitizing operation via the **ZVL:** command with no effect on other fields in the digitizer or format. The fifth field is the event field and is identified by an "E." This field contains five spaces but should not be filled in excess of 32 000 points. The sixth field is a free text, alphanumeric area consisting of four groups of eight spaces. This field can be accessed at any time during the digitizing operation via the **DAT: DAT1: DAT2: DAT3:** or **DAT4:** command with no effect on other fields in the digitizer or format. Therefore, the most flexible format relative to data collection via the digitizer is 1F, 10X, 10Y, 8Z, 5E, 32C. After completion of step 6 in the previous section, enter **FMT: 1F, 10X, 10Y, 8Z, 5E, 32C** and press the return key. If it is suspected that the format is lost during a digitizing operation, then it is necessary to reload the format. It should be noted that this format will place good data on the tape when the UTM grid system is used. The digitizer simply stores data on tape in a prescribed format. Further processing of this data must be accomplished with a system such as that described in references 1 and 2.

Creation of Data File

In the section "Creation of Tape Files," a file number was assigned, depending on what situation was encountered. Data were taken during the scale factor computation and stored in the buffer. It is now necessary to clear the buffer and store good data in the file established in the "Creation of Tape Files" section.

Step 1. Enter **CLS** and press the return key. This command will write to tape all data contained in buffer and place an end-of-file marker following the data.

Step 2. Enter **MTP:** and the file number used in the section on creation of tape files. After nine data points have been taken, the digitizer will rewind the tape to the load point and count the end-of-file markers until one number lower than the one specified is sensed. Data will then be stored in the next file until a **CLS** command is given.

Documentation

Use the **TXT:** command to enter one line of alphanumeric data. The following is a minimum: format used, coordinates and scale factors used at rotation and initialization, project name, job order number, operator's name, and current date. Other data deemed necessary can be entered in this manner.

Creation of Variable Data

There are three areas of variable data in the systems output. They are (1) the "Z" string, consisting of eight spaces of alphanumeric data accessed by the **ZVL:** command; (2) the "C" string, consisting of 32 spaces of alphanumeric data in four groups of eight and accessed by **DAT:** (all 32 spaces), **DAT1:** (first eight space group), **DAT2:** (second eight space group), **DAT3:** (third eight space group), and **DAT4:** (fourth eight space group); and (3) a **TXT** string, consisting of 62 spaces of alphanumeric data and accessed by the **TXT** command.

Data entered via the **ZVL** and **DAT** commands are printed in each data point. Data entered via the **TXT** command is printed directly on the tape one time. Information can be entered into these areas at any time without affecting the format or digitizing operation. Because these areas are free text, it is advisable to establish what data to enter in each area prior to initiation of the digitizing task.

Digitizing Point Data

This section is devoted to the point mode for digitizing data.

Step 1. Locate the point/line mode switch under the top of the back-lighted table (fig. 4), and move upward to the point position.

Step 2. Enter **EVT: 1** and press the return key.

Step 3. Enter **CVCT** and press the return key.

Step 4. Begin recording data, using the variable data areas, to identify data points or a group of data points.

Step 5. When an area of logically grouped points has been completed, enter **CLS** and press the return key.

Step 6. Enter **MTP:** and the next higher file number, and continue to record data. As a minimum, close the file when moving from one source sheet to another.

Step 7. When the digitizing is complete for a given tape, enter **CLS** and press the return key. If two consecutive end-of-file markers are not placed after the last data file, the processor reading the tape will go into runaway operation after the last file. After two **CLS** commands are executed, go to the section "Tape Demounting" and continue.

Digitizing Linear Data

Prior analysis and organization of the linear data digitizing requirement is essential if accurate, complete, and timely data are to be recorded. Since the nature of the system dictates an item-by-item, sheet-by-sheet, frame-by-frame, or piece-by-piece method of digitizing, source material analysis and organization should follow steps 1 to 10.

Step 1. Analyze the source on which linear features will be digitized and break all linear features over 10.16 centimeters (4 inches) in length into approximate 10.16-centimeter (4 inch) segments. For all features that are nonlinear, establish a point on the feature that will be the start-and-stop point. Use a fine-point pencil or pen when making these marks. The overall accuracy of the data depends on careful attention to detail and thorough completion of the tasks.

Step 2. Starting at the upper left corner of the source material, number each linear segment or nonlinear feature in ascending order and in approximately 10.16-centimeter (4 inch) bands from left to right until the entire source has been numbered. Steps 1 and 2 can be done either on the digitizer table or elsewhere. If they are not done on the digitizer table, it is assumed that the foregoing procedures have been followed through the previous section.

Step 3. Locate the point/line mode switch (fig. 4) and move the switch downward to the line position.

Step 4. Locate the rate switch (fig. 4) and rotate fully clockwise.

Step 5. Enter **VCT:** and the distance between data points decided upon by the principal investigator as a nondecimal figure. Press the return key.

Step 6. Enter **EVT:1** and press the return key.

Step 7. Enter **ZVL:** and the number assigned in step 2.

Step 8. Place the cursor at the start point or left end of item 1, as numbered in step 2. Press cursor button 1, 2, 3, or 4, and carefully follow the feature to the stop point or to the end of the segment. Release the cursor button, and draw a line through the number. If at any time while taking data in the line mode a beep is heard from the processor, then data have been taken faster than the computer can perform the necessary calculation, and data have been lost. Go to step 7 and reaccomplish this segment or feature.

Step 9. Continue this sequence until data intended for this file are complete. Enter **CLS** and press the return key.

Step 10. Enter **MTP:** and next higher file number. Press the return key. If data will be taken from the source material that is mounted, scaled, and rotated, setup variable data areas and continue taking data. If the source material is not mounted, scaled, and rotated, go to the section "Creation of Tape Files" and continue.

Step 11. When the digitizing has been completed for a given tape, enter **CLS**. Press the return key twice and go to the section "Tape Demounting." If two consecutive end-of-file markers are not present after the last file on the tape, the computer processing of the data will go into runaway operation.

CONCLUDING REMARKS

The NASA Earth Resources Laboratory (ERL) has defined and verified, on an operational basis, a procedure that uses a semiautomated digitizer system for converting disparate sources of data contained on maps, charts, and film into a digital format for further processing on a mainframe computer system. The various digitizer system components, assembled for the ERL by the Altek Corporation, are an integral part of a computer system that has evolved at the ERL through previous research efforts. These various digitizer system components have been described. In addition, the detailed system operating procedures consisting of functions, commands, and steps are presented in a simplified and logical manner and are directly applicable to most digitizing applications.

Several graphic digitizing considerations, as they relate to ERL's Digitizer System, are presented for use by organizations that are contemplating the use of map-based data to assist them in deriving information upon which to make timely and critical management decisions. These considerations include computer processing, geographic map sources, characteristics, and classifications. Emphasis is also placed on the various types of digitizers, data characteristics, and data structuring methods to consider in digitizing applications.

Earth Resources Laboratory, NSTL
National Aeronautics and Space Administration
NSTL Station, Mississippi March 1979
177-52-85-TA

REFERENCES

1. Whitley, Sidney L.: A Procedure for Automated Land Use Mapping Using Remotely Sensed Multispectral Scanner Data. NASA TR R-434, 1975.

2. Whitley, Sidney L.: Low-Cost Data Analysis Systems for Processing Multispectral Scanner Data. NASA TR R-467, 1976.

Appendix A: Geographic Map Sources, Characteristics, and Classifications

SOURCES

The National Cartographic Information Center (NCIC) of the U.S. Geological Survey, Department of the Interior, provides a national information service to make cartographic data of the United States accessible to the public and to various federal, state, and local agencies (ref. A-1). These cartographic data include geographical maps and charts, aerial and space imagery, soil maps, topographic maps, etc. Basically, these data are a fundamental source for the acquisition, measurement, analysis, storage, inventory, display, and dissemination of information concerning the environment and its resources and, as such, are an important consideration of planners of systems for producing useful management information. This type of data is typically used by the NASA Earth Resources Laboratory (ERL) as its source of topographic information. It exists in computer-compatible tape format. Several programs exist within the ERL that provide the capability to analyze these topographic data. One program reformats the raw, digital terrain data as obtained from the NCIC. Another program converts these data to slope, slope length, aspect, etc. These data can be properly formatted for entry into a data base and read into the ERL image display system and displayed as a color-coded map on a TV screen.

CHARACTERISTICS

The multiuse geographical maps are a representation on a plane surface, at any appropriate scale and projection, of a portion of the Earth's surface.

These maps contain symbols and textual data to identify features and thematic information. Topographic maps, for example, use line-and-symbol representations of natural and selected man-made features of a part of the Earth's surface plotted to a definite scale (ref. A-2). A characteristic of this map is the portrayal of the shape and elevation of the terrain by contour lines. The scale of these maps ranges from the standard 7.5 × 7.5 minutes quadrangle series at a large scale of 1:24 000 to those series at a smaller scale of 1:1 000 000. The 1:250 000-scale maps published by the U.S. Geological Survey are drawn on a transverse Mercator projection and are generally in quadrangle units of 1° latitude by 2° longitude (ref. A-3). Within ERL, reference A-4 contains information concerning an application example used in keying Landsat-data-derived classifications of the 1:250 000-scale series maps prepared by the U.S. Geological Survey, Department of Interior. Points, areas, heights, and distances can be extracted from these maps and used directly in application programs or input to a data base for further use.

The relationship of the geographical map to the figure of the Earth is another important characteristic. Maps that are to be used for analyzing relationships require conformal projections. (See reference A-5, page 221 for a note on "conformality.") Some of the conformal projections are (1) Mercator's, (2) Lambert's Conic, and (3) Stereographic. The most widely used are Lambert's Conic and a transverse form of the Mercator. This latter projection has been widely used to provide a projection base for topographic maps. The Universal Transverse Mercator (UTM) projection uses a square grid that is superimposed on the map. The UTM grid is used to represent areas of the Earth between the 80° north and south parallels of latitude that coincide with east-west direction. These parallels of latitude are divided into 60 north-south columns or zones of 6 degrees of longitude wide. Each zone contains a central meridian that is given an easting value of 500 000 meters. The Equator is given an arbitrary value of 10 000 000 meters for the Southern Hemisphere. Distances measured in meters west to east are designated "eastings" and from south to north are designated "northings." The grid interval is normally at 10 000-meter intervals.

This type of reference system is used quite frequently within the ERL in dealing with coastal and land applications wherein data management type software is required. Thus, the majority of the digitizing requirements within the ERL is concerned with the use of maps that are based on the UTM reference grid system.

CLASSIFICATIONS

The various types of map classification categories are point data, comprehensive line, isovalue, and comprehensive area (ref. A-6). The point data portrays point location of a single attribute. A characteristic of the line data is that any point on a line has a single attribute and two or more adjacent points of the same attribute form a line segment. On an isovalue type classification, the attributes are fixed incremental values of a particular descriptor (i.e., topographic contour, isotherm, etc.). A given line would portray a locus of equal descriptor values. The comprehensive area classification exhibits area attributes wherein each location within a boundary corresponds to a specific type area.

All these type-classification maps are used to some degree within the ERL. Perhaps the most frequently used is the comprehensive area followed by the point data and comprehensive line.

REFERENCES

A-1. National Cartographic Information Center (NCIC) User Guide. U.S. Department of the Interior, Geological Survey, INF-74-47.

A-2. Topographic Maps. U.S. Department of Interior, Geological Survey, Information Pamphlet 024-001-02793-3, 1977.

A-3. Index to Topographic Maps—1:250 000-Scale Series. U.S. Department of Interior, Geological Survey, July 1976.

A-4. Graham, M. H.: Digital Overlaying of the Universal Transverse Mercator Grid With Landsat-Data-Derived Products. NASA TM 58200, 1977.

A-5. Robinson, A. H.; and Sale, Randall D.: Elements of Cartography. Third ed., John Wiley & Sons, Inc., 1969.

A-6. The Development of a Natural Resource Information System. Raytheon Co., Wayland, Mass., Vols. I and II, July 1973.

Appendix B: Graphic Digitizing

DISCUSSION

A digitizer is a device used for converting data or information from a source graphic map to a digital format suitable for computer processing. Graphic digitizing is the translation or conversion of spatially distributed information (i.e., lines or points on a map) into one or more series of numeric values. Basically, what occurs in this process is that the position of a stylus or cursor on a graphic surface is converted into a digital output. Keyboards are attached to most digitizers to allow for operator interaction in entering labels, codes, or other information associated with the graphical representation. In addition, the operator uses the keyboard to enter commands that perform programed instructions.

This conversion of graphic data into an appropriate digital computer format is probably the most time-consuming operation involved in providing input data to an information management type computer-based system.

TYPES OF DIGITIZING EQUIPMENT

The equipment that is commercially available for converting graphic map-type data into a digital format can be classified as being manual or automatic. However, as will be pointed out later, the fully "automated" system requires certain manual procedures to be performed, and it is approximately 10 times more expensive than a comparable manual system that could perform the same digitizing task.

Basically, the different types of digitizers are summarized in reference B-1 and are listed in the following chart:

Type digitizer	*Characteristics*
Line	Records one point at a time.
Continuous time sampling	Records points at a set time interval.
Continuous distance sampling	Records points at threshold values.
Automatic line following	Uses electro-optical scanning techniques.

This list is not all inclusive but is intended to be an indicator of the general types of available equipment. The system in use at the NASA Earth Resources Laboratory (ERL) has the capability to record one point at a time or to record points at threshold values.

The line digitizer generates either point or line data. It consists of a flat table for mounting the source input map. The table can be adjustable or fixed and generally is back-lighted. The sensor can be a movable head with a five-tip pointer connected to a position-sensing instrument. It could also be a cursor in the form of a movable tablet-type device that has fine-line crosshairs located in a transparent miniglass window. When positioned on the grid surface, and when the appropriate key on the cursor is pressed, these crosshairs activate a magnetic grid in the surface of the digitizer table and capture the desired point data.

If an entire line is to be captured, the cursor is traced manually along the entire line, and the line is automatically converted into a series of X-Y coordinates. These converted points can be either equally incremented in the vertical or horizontal direction (called continuous distance sampling) or recorded at an equal time interval (called continuous time sampling).

The line digitizer is the most common type of manual equipment. It captures and either displays or stores in some storage media (i.e., magnetic tape, memory, etc.) the position of the cursor's distance from some predetermined origin or reference point. In the point mode, a cursor position is normally determined when a key on the cursor is pressed. This type of digitizing allows the most interaction on the part of the operator and usually requires supporting input-type labeling information by the operator. In some cases, however, it is extremely slow from the standpoint of data capture rate because of the dependence on the manual labeling process.

The continuous distance-sampling digitizer has the capability to record coordinate values only when the X-Y coordinate values being monitored exceed a selected threshold distance. Thus, a prefixed space resolution between any two consecutive digitized points is maintained. On the other hand, the continuous time-sampling digitizer records the position of the cursor at a set time interval that is generally an operator input parameter. This type of digitizer usually requires a minicomputer with a magnetic tape recorder to keep up with the high rate and volume of samples that are normally captured. This type of digitizing allows the capture of more points with small increments along a line or curve that has a sharp change or large curvature. In this situation, the rate of movement of the cursor is slower than usual. Many digitizers have the capability for both distance- and time-sampling modes of operation. The ERL system has the continuous distance-sampling capability as well as the line capture capability (one point at a time). These systems, which include the ERL system, are particularly suited for polygon digitizing in which graphic detail is recorded directly and boundary detail is preserved.

The automatic line-following digitizer is an electro-optic unit that bisects the width of the graphic line and uses additional circuitry to automatically follow the line being digitized with accurate positioning at the midpoint of the line width.[2] The lines used in this process must be high contrast and continuous. This in itself would require a retracing operation to be performed on the area under consideration. The time and cost associated with preparing this retrace would be almost the same amount of effort required for digitizing directly from the original maps by one of the previously mentioned digitizers. One such automatic line-following system (LASCO) is discussed in an NASA NSTL internal report by Su and Montgomery.[2] The typical cost ($500 000 to $700 000) for a system such as the LASCO is approximately 10 times more expensive than conventional stand-alone systems such as the one used at the ERL. It should be noted that the ERL digitizer system is not of the automatic line-following type. Additional information pertaining to a test case using the LASCO optical line-following digitizer is contained in the previously mentioned report.

DATA CHARACTERISTICS

The characteristics of the converted data have certain accuracy, repeatability, and density requirements associated with the particular digitizing equipment and with the application. Individual data points must have order or connectivity and thus must be coded or labeled and stored as a chain of discrete points. The accuracy is associated with the operator's ability to position the stylus on a unique point. Density has to do with the interval between consecutively digitized points. Repeatability is associated with the hardware system in the sense that if the operator comes back to a previously digitized point and recaptures this point, then, all other variables being the same as before, this discrete point will be repeatable within plus or minus limits. This repeatability characteristic usually varies from one system to another.

DATA STRUCTURING METHODS

There are basically three methods that can be used to convert graphic data (i.e., maps) to digital computer files (ref. B-1). These are the grid cell, polygon, and line segment methods. These three methods are shown in figure B-1. The grid cell method divides a map into rectangular grids, and information within a grid is stored in a matrix-type array (i.e., rows and columns). Proper coding for each cell is required to tag each cell with its predominant attribute. The polygon method uses X-Y coordinates that describe each individual area with respect to a particular attribute. The line segment method uses X-Y coordinates from point to point along boundary attributes and is somewhat similar to the polygon approach. All these methods have their advantages and disadvantages. The selection of one over the other would, to a great extent, depend on the particular application. All three methods are used to some degree within the ERL.

One advantage of the grid cell method is that grid-cell formatted digital data for a particular attribute can be combined with grid-cell data for another attribute to produce an output tape that contains the different attribute information over the same area. One obvious disadvantage to the grid cell method is that the map reproduced from the digital data may show a step-type line along boundary areas representing different attributes. This is due to the selection logic that determines if a cell is or is not assigned a particular attribute; i.e., if the center point of a cell is outside and above an attribute area, then the cell is tagged to be void of

[2] An NASA NSTL internal report entitled "Cartographic Map Digitization" has been prepared by M. Y. Su and F. L. Montgomery (ESSL-75-009), 1975.

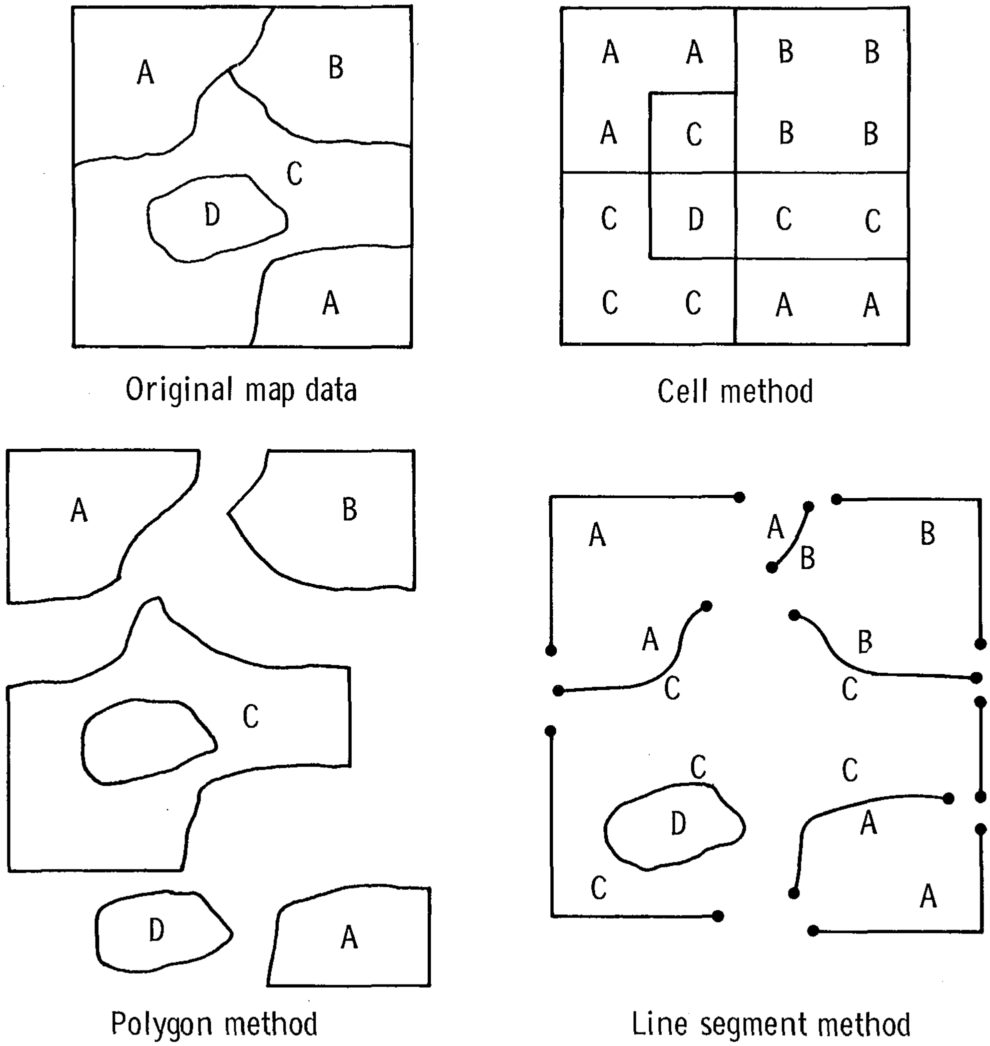


FIGURE B-1.—Data structuring methods.

this particular attribute. The selection logic could also be based on the percentage of the cell occupied by the particular attribute.

In the polygon approach, the graphic detail is recorded directly, and the boundary detail can be preserved. The digitized data consist of ordered pairs of coordinates giving line segments or boundaries around each attribute. This type of approach generally requires less storage on tape or disk. For users whose primary interest is storing and graphically displaying raw data, this is a very accurate and efficient approach. In the line segment method, all the line detail is encoded as a series of X-Y coordinate values. A series of values terminates where two or more other lines terminate. These coordinate values are reformatted such that for a given series or line segment there are assigned arbitrary left and right region numbers. In addition, the attributes on the left and right of the line segment are given some type of designation. The advantage of this method is that it can be used to reproduce point and line features of the original source map. The disadvantage of this method is that it is difficult to combine attribute data with other types of attribute data over the same area.

REFERENCE

B-1. The Development of a Natural Resource Information System. Raytheon Co., Wayland, Mass., Vols. I and II, July 1973.

1. Report No. NASA RP-1048	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Procedure for Extraction of Disparate Data From Maps Into Computerized Data Bases		5. Report Date October 1979
		6. Performing Organization Code
7. Author(s) Bobby G. Junkin		8. Performing Organization Report No. S-491
9. Performing Organization Name and Address Earth Resources Laboratory* NSTL Station, Mississippi 39529		10. Work Unit No. 177-52-85-TA
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Reference Publication
		14. Sponsoring Agency Code

15. Supplementary Notes

*The Earth Resources Laboratory is an element of the National Space Technology Laboratories, NSTL Station, Mississippi 39529.

16. Abstract

A procedure is presented for extracting disparate sources of data from geographic maps and for the conversion of these data into a suitable format for processing on a computer-oriented information system. Several graphic digitizing considerations are included and related to the NASA Earth Resources Laboratory's Digitizer System. Current operating procedures for the Digitizer System are given in a simplified and logical manner. The report serves as a guide to those organizations interested in converting map-based data by using a comparable map digitizing system.

17. Key Words (Suggested by Author(s))		18. Distribution Statement
Information systems	Mapping	STAR Subject Category
Earth Resources Laboratory	Imagery	60 (Computer Operations and Hardware)
Computerized map data	Digital format	
Computer graphics	Map digitizing	
Data conversion	Land use maps	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	23	$3.50

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1979

National Aeronautics and
Space Administration

Washington, D.C.
20546

Official Business
Penalty for Private Use, $300

THIRD-CLASS BULK RATE

Postage and Fees Paid
National Aeronautics and
Space Administration
NASA-451

U.S.MAIL

NASA

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return